DOCUMENT RESUME

ED 181 423 CS 005 200

TITLE Testing and Evaluation in Reading and Communication Skills: Abstracts of Doctoral Dissertations Published in "Dissertation Abstracts International," July through December 1979 (Vol. 40 Nos. 1 through 6).

INSTITUTION ERIC Clearinghouse on Reading and Communication Skills, Urbana, Ill.

PUB DATE 79

NOTE 16p.

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Academic Achievement; Adult Basic Education; Adults; Communication (Thought Transfer); *Composition (Literary); Doctoral Theses; *Educational Research; Elementary Secondary Education; *Evaluation; Higher Education; Minimum Competency Testing; Predictive Measurement; Readability; Reading Instruction; *Reading Research; *Reading Tests; Speeches; Student Placement; *Testing; Test Validity

IDENTIFIERS *Writing Evaluation

ABSTRACT

This collection of abstracts is part of a continuing series providing information on recent doctoral dissertations. The 20 titles deal with a variety of topics, including the following: predictors of primary grade reading achievement; factor analysis of the WISC-R to distinguish disabled readers at three grade levels; a process model for developing informal oral reading inventories; the curricular validity of selected standardized reading achievement tests; an investigation of speech ratings as a function of rating scale; a comparison of approaches to measuring outcomes in adult basic education; the relationship between high school reading and math achievement levels and later life experiences; a correlation analysis of local and state minimum competency standards in New Jersey schools; teacher expectancy--the effect of race, sex, direction of writing performance and trials on the grading of essays; placement in freshman composition classes; and a factorial analysis of essay evaluation. (HKM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED181423

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Testing and Evaluation in Reading and Communication Skills:

Abstracts of Doctoral Dissertations Published in Dissertation Abstracts International, July through December 1979 (Vol. 40 Nos. 1 through 6)

Compiled by the staff of the

ERIC Clearinghouse on Reading and Communication Skills

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

University Microfilms International

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS005200

The dissertation titles contained here are published with permission of the University Microfilms International, publishers of Dissertation Abstracts International (copyright © 1979 by University Microfilms International), and may not be reproduced without their prior permission.

This bibliography has been compiled as part of a continuing series designed to make information on relevant dissertations available to users of the ERIC system. Monthly issues of Dissertation Abstracts International are reviewed in order to compile abstracts of dissertations on related topics, which thus become accessible in searches of the ERIC data base. Ordering information for the dissertations themselves is included at the end of the bibliography.

Abstracts of the following dissertations are included in this collection:

Almenoff, Phyllis
A COMPARISON OF KINDERGARTEN PREDICTORS FOR FORECASTING SECOND GRADE READING ACHIEVEMENT

Applegate, Anthony Joseph
THE CONSTRUCTION AND VALIDATION OF A TEST OF READING COMPREHENSION AT THE COLLEGE LEVEL

Brooks, Harry F.
THE CURRICULAR VALIDITY OF SELECTED STANDARDIZED READING ACHIEVEMENT TESTS

Brothers, Emily Severance
THE EFFECT OF THREE TYPES OF TEACHER EVALUATION ON THE COMPOSITIONS OF COLLEGE FRESHMEN

Cervero, Ronald Michael
A COMPARISON OF APPROACHES TO MEASURING OUTCOMES IN ADULT BASIC EDUCATION

Cramer, Genevieve Rummell
AN EXAMINATION OF THE VALIDITY AND UTILITY OF A LONG WORD MEASURE IN THE PREDICTION OF READABILITY

Edmunds, Robert French
AN INVESTIGATION OF SPEECH RATINGS AS A FUNCTION OF RATING SCALE FORM, EVALUATOR TYPE, AND TRIALS

Furr, Jane Dellana
THE RELATIONSHIP BETWEEN HIGH SCHOOL READING AND MATH ACHIEVEMENT LEVELS AND LATER LIFE EXPERIENCES: IMPLICATIONS FOR MINIMAL COMPETENCY TESTING

Grove, Martha Judin
SAMPLE FREQUENCY MODIFICATIONS AND COMPARATIVE READABILITY OF FOUR READABILITY FORMULAS AS APPLIED TO SELECTED TEXTBOOKS ON THE TEACHING OF READING

Johnson, Sally Hartman
THE DEVELOPMENT OF AN INSTRUMENT TO MEASURE HUMAN COMMUNICATION FUNCTIONAL DOMINANCE

Leimgruebler, Zoe Walker
THE ESTABLISHMENT OF RELEVANT SIXTH-GRADE NORMS ON A READING TEST CONSTRUCTED FOR GRADES 7-9

Lofgren, John O.
THE DEVELOPMENT AND VALIDATION OF THE PRESCHOOL LANGUAGE-COGNITIVE SKILLS ASSESSMENT FOR CURRICULUM ENTRY

Nadler, Frederick Francis
A CORRELATION ANALYSIS OF LOCAL AND STATE MINIMUM COMPETENCY STANDARDS IN NEW JERSEY SCHOOLS

Porter, Patricia Thomas
TEACHER EXPECTANCY: THE EFFECT OF RACE, SEX, DIRECTION OF WRITING PERFORMANCE AND TRIALS ON THE GRADING OF ESSAYS

Rentz, Sarah Taylor
PRE-EMPLOYMENT TEST SCORES AS PREDICTORS IN A REMEDIAL READING PROGRAM

Shale, Douglas Gordon
A FACTORIAL ANALYSIS OF ESSAY EVALUATIONS

Smith, Carolyn Crawford
A DESCRIPTIVE STUDY OF COMPETENCY BASED EDUCATION AS RELATED TO ACHIEVEMENT, READING ABILITY AND GRADE POINT AVERAGE OF HIGH SCHOOL STUDENTS

Smith, Merrill Lee
THE USE OF KAUFMAN'S FACTOR ANALYSIS OF THE WISC-R IN DISTINGUISHING ABLE AND DISABLED READERS AT THREE GRADE LEVELS

Stenberg, Daniel Raymond
A COMPARISON OF ERRORS MADE ON THE WEEPT AND THOSE FOUND IN STUDENT WRITING SAMPLES AS BASES FOR PLACEMENT IN FRESHMAN COMPOSITION CLASSES

Watkins, William C.
THE CONSTRUCTION AND VALIDATION OF A PROCESS MODEL FOR DEVELOPING INFORMAL ORAL READING INVENTORIES

A COMPARISON OF KINDERGARTEN PREDICTORS FOR FORECASTING SECOND GRADE READING ACHIEVEMENT
Order No. 7918421

ALMENOFF, Phyllis, Ed.D. Hofstra University, 1979. 234pp.

The major purpose of this study was to assess the predictive efficiency of selected readiness measures to determine how well they forecasted reading achievement in second grade when assessed with a standardized reading test and a cloze test. This study also sought to determine the best predictors for identifying students who have a high risk of failure and for identifying gifted readers.

The subjects for this three-year retrospective study were 277 second grade students on whom complete kindergarten and second grade testing data were available. The students came from nine elementary schools in Baldwin, New York, a suburban school district. There were 150 boys and 127 girls in the study.

Kindergarten testing included the following standardized tests: (1) the Vane Kindergarten Test, (2) the Boehm Test of Basic Concepts, and (3) the Metropolitan Readiness Test (MRT). In addition, teachers rated the students using the Kindergarten Profile of Developmental Characteristics (an instrument developed by the school district) which assessed (1) perceptual-motor development, (2) social-emotional readiness, (3) maturational characteristics, and (4) language and speech patterns. The kindergarten testing and ratings were compared with second grade performance on the Comprehensive Test of Basic Skills (CTBS) and a cloze test administered by the investigator. Chronological age and sex were also assessed as predictors.

The data were analyzed using Pearson product-moment coefficient of correlation, multiple regression formulas, and factor analysis.

Findings

1. The best kindergarten predictors of second grade reading achievement as measured by the CTBS were (a) the MRT, (b) the Vocabulary subtest of the Vane Kindergarten Test, and (c) self-concept.
2. The most efficient predictors of second grade reading achievement as measured by the cloze test were (a) the Number subtest of the MRT, (b) the Vocabulary subtest of the Vane Kindergarten Test, and (c) the Word Meaning subtest of the MRT.
3. There was not a significant difference in the efficiency of kindergarten predictor variables when second grade reading achievement was measured by (a) a standardized reading test and (b) a cloze test.
4. High risk students were seen as a different group by the kindergarten predictor variables. The best predictors of low reading achievement were (a) language structure, (b) the Numbers subtest of the MRT, and (c) ease.
5. Students who were gifted readers were not seen as a different group by the kindergarten predictor variables.
6. Chronological age and sex did not correlate significantly with reading achievement at the end of second grade.
7. Factor analysis extracted a cognition composite from the kindergarten variables.
8. A combination of kindergarten variables added significantly to the value of the MRT in predicting second grade reading achievement.
9. A combination of kindergarten predictors explained (a) 45 percent of the variance on the CTBS and (b) 43 percent of the variance on the cloze test.

Conclusions

1. Many of the kindergarten tests and ratings measure similar skills.
2. Although the Metropolitan Readiness Test was the best predictor of second grade reading achievement, the use of other tests and ratings added significantly to the efficiency of the prediction. Nonetheless, a great deal of the prediction remained unexplained.
3. Cognition was a major factor in successful reading achievement.
4. The sex and chronological age of a subject did not have significant prognostic value.
5. The same factors which predict reading achievement for all second grade students are effective for predicting gifted readers.
6. A different set of predictors should be used to predict students who have a high risk of low reading achievement.
7. Perceptual-motor factors had very little prognostic value.
8. Predictors for reading achievement were not very different whether assessed by a standardized test or a cloze test.
9. Reading is an extremely complex skill requiring linguistic ability.

THE CONSTRUCTION AND VALIDATION OF A TEST OF READING COMPREHENSION AT THE COLLEGE LEVEL
Order No. 7923978

APPLEGATE, Anthony Joseph, Ph.D. Temple University, 1978. 145pp.

The purpose of this study was the construction and validation of a test of reading comprehension designed to measure three types of reading ability among college students. In short, the test was designed to differentiate among students who 1) can comprehend only the literal message of the author (Type A reading), 2) can also translate that message into different words or different forms (Type B reading), and 3) can interpret the factual material they have comprehended in such a way as to arrive at conclusions or implications not directly stated in the communication (Type C reading).

The Experimental Form of the reading comprehension test consisted of 15 selections written by the experimenter and ranging in length from 140 to 240 words. Each selection was accompanied by 10 multiple choice format test items, with 3 as Type B and 4 as Type C. Five selections were included at each of the following Dale-Chall readability level ranges: 5th-6th; 7th-8th; 9th-10th. Thus the Experimental Form of the test included 15 Type A, 15 Type B and 20 Type C items at each of the three readability ranges for a total of 150 items. In this form, the test yielded scores on the nine item type x readability level subtests, 3 item type subscales (totaled across readability levels), and the total test. The Experimental Form of the test was subjected to critical review by eight reading specialists and administered to a sample of 75 students from three area colleges. As a result of these procedures, the test was thoroughly revised.

The Revised Experimental Form (REF) of the test was administered to 201 students from five area colleges and the results were analyzed to determine the extent to which the test could 1) reliably measure Type A, B and C reading skills and 2) distinguish between Type A, B and C skills in its measurement.

The reliability coefficients for the nine subtests of the REF ranged from .54 to .70. However, since the readability level of the passages had no consistent effect on the performance of students, the item type subscales proved to be of greater interest. The reliability coefficients of the Type A, B and C subscales were .85, .84, and .82, respectively, and the reliability for the total 150 item test was .93.

As a measure of the REF's ability to distinguish among groups of reading skills, the subtests and subscales were correlated and the correlations corrected for attenuation. The corrected correlation between the Type A and Type C subscales was .77, while the correlation between Type B and Type C approached 1.00. Thus the REF appeared to provide a relatively distinct measure of two rather than three types of skills. This finding was supported by the results of a factor analysis of subtest scores. Two significant factors were extracted, with Type B and Type C subtests loading highly on Factor I and Type A subtests loading highly only on Factor II.

It was concluded that the REF provided a relatively distinct measure of literal (Type A) vs. inferential (Types B and C) reading skills, but could not distinguish between Type B and

Type C in its present form. An analysis of item-subscale correlations shed considerable light on the structure of Type B and Type C abilities and led to the conclusion that a revised and shortened form of the REF could widen the distinction in measurement between the two sets of abilities. It was suggested that if such a revision could be administered to a much larger sample, a factor analysis of items might provide useful information in the construction of a test that could distinctly measure three rather than two groups of reading skills.

THE CURRICULAR VALIDITY OF SELECTED STANDARDIZED READING ACHIEVEMENT TESTS Order No. 7918229

BROOKS, Harry F., Ed.D. State University of New York at Albany, 1979. 213pp.

This study was conducted in an attempt to establish the curricular validity of standardized reading achievement tests for a representative sample of New York State pupil school districts. The sample, stratified by geographical region, consisted of 145 districts.

It was assumed that a valid index of curricular validity was the congruence that exists between the relative emphasis of content categories of district curriculum and the relative emphasis of categories of standardized achievement test content.

The determination of curricular validity of the tests studied (i.e., the Metropolitan Achievement Test, the Iowa Tests of Basic Skills, and the New York State Pupil Evaluation Program Reading Test) entailed three major components. These were: (1) specifying district reading curriculum, (2) classifying test content into curriculum categories, and (3) analyzing the relationship between curriculum content and test content.

A problem of secondary importance in this study was an analysis of the appropriateness of test passage reading difficulty. The relationships between district characteristics and curricular validity were also analyzed.

The following conclusions were tentatively drawn from this study:

(1) The content of standardized reading achievement tests demonstrated a substantial lack of congruence to the curriculum content of New York State public schools.

(2) Overall, districts of New York State did not utilize the standardized reading achievement test which provided the greatest curricular validity for their reading programs.

(3) The content difficulty levels of commonly used standardized reading achievement tests fell within district representatives' perceptions of appropriate difficulty ranges for their students.

(4) Districts of New York State utilized the test with greatest difficulty level appropriateness at the third grade level.

(5) There was no significant relationship between district per pupil expenditure for instruction and curricular validity. There was no significant relationship between time allocated for reading instruction and curricular validity of standardized reading tests.

(6) The proportion of district student populations classified as nonracial-ethnic was related in a positive direction to curricular validity.

(7) There was a positive relationship between district achievement on the PEP test and the curricular validity of that test.

It was recommended that educators conduct local analyses of standardized test content to establish the curricular validity of their evaluation instruments. It was recommended that future studies verify the methodology of this study.

THE EFFECT OF THREE TYPES OF TEACHER EVALUATION ON THE COMPOSITIONS OF COLLEGE FRESHMEN Order No. 7920026

BROTHERS, Emily Severance, Ph.D. University of South Carolina, 1979. 186pp.

Determining the relative effectiveness of three types of teacher evaluation on the written compositions of low-ranking college freshmen, comparing analytic and holistic scoring of paragraphs, and finding student preferences for these methods of evaluation were the purposes of this study. The three methods of evaluation included negative comments (negative markings and comments only), positive comments (a few negative markings but mainly positive markings and comments), and conference (positive markings and a student-teacher conference to correct the errors).

Having made fewer than forty-five of a possible sixty points on the Test of Standard Written English (TSWE),[1] 180 college freshmen spent two extra hours a week in a writing laboratory following a programmed text. On the basis of low, medium, and high reading levels on The Nelson-Denny Reading Test,[2] the subjects were randomly assigned to the three methods groups with equal numbers of students from the three reading levels. During the fall term of 1978–1979, these subjects wrote in the writing laboratory under controlled conditions paragraphs on uniform topics of general interest. From 300 to 500 words in length, these paragraphs called for thesis statements, three or more points of proof, and summary sentences. In preparation for writing, students were permitted to discuss the assigned topics for five minutes before writing for the remainder of the hour.

Trained and rehearsed in marking by the three evaluation methods, five English Department faculty members marked the themes the particular week of the writing. Students could therefore see their returned papers before writing their next ones and could use the markings as guides in improving their new paragraphs. The writing continued for a total of five experimental paragraphs. On the last writing day, students responded to a Student Questionnaire, reacting to their experiences.

For the holistic judging of themes, the English faculty devised and used a scale based on control of thought, use of language, and mechanics: A or 5; B or 4; 3 or C; 2 or B; and 1 or F. The investigator designed an analytic scale evaluating organization, mechanics, and sentence structure on a basis of twenty-five points to be divided by five to match the holistic scale. After inservice sessions and practicing, five faculty members judged the papers holistically with the investigator doing the recording; five others judged them analytically with the investigator recording. The means of these analytic and holistic scores were used in determining the findings in the study.

For comparison of means of the three methods of evaluating student compositions, the analyses of variance showed no significant differences at the 0.1 level of confidence. Students, however, made mean score gains from theme 1 to theme 5 of 0.72 of a grade point by analytic scoring and of 0.17 by holistic scoring. For the comparisons of means between analytic and holistic scoring, no significant differences were found. For student preferences for methods of evaluation, 54.44 percent of the students preferred the conference method; 21.11, positive comment; 20.00, negative comment; and 4.44, no method of any kind.

Experimentation in methods of evaluation extending over a longer time period than one semester should prove profitable. Students should have access to writing teachers for conferences and should help make decisions as to kinds of evaluations. English Departments should examine their methods of evaluating student writing and try constantly to improve these methods.

[1] College Board Scholastic Aptitude Test, Princeton: College Examination Board, Educational Testing Service, 1971.

[2] M. J. Nelson and E. C. Denny, The Nelson-Denny Reading Test: Vocabulary, Comprehension, Rate, Atlanta: Houghton Mifflin Company, 1973.

A COMPARISON OF APPROACHES TO MEASURING OUTCOMES IN ADULT BASIC EDUCATION

CERVERO, Ronald Michael, Ph.D. The University of Chicago, 1979

While each state has a system of publicly funded adult basic education (ABE) programs, a fundamental concern has been raised that the content taught and the methods used in these programs have been borrowed from the elementary and secondary educational systems and, therefore, are inappropriate for use in ABE. As a result, a new approach, adult performance level (APL), to defining, teaching, and measuring knowledge and skills in ABE was developed and is being implemented with no empirical evidence to indicate that this approach significantly increases student enrollment, retention or achievement in ABE programs.

The purpose of this study was to determine how the APL approach to defining and measuring outcomes in ABE compares with the most widely used approach to defining and measuring outcomes exemplified by the General Educational Development (GED) test. It was hypothesized that the APL approach, which is designed to measure essential life coping skills, and the GED test, which is designed to measure the long term outcomes of a high school education would be highly positively related since both tests purport to measure reading, writing and computation. Further, the validity of two versions of the APL test, the Texas APL test and the APL Survey, were examined. The APL Survey was administered to 296 adults at the time of the administration of the GED test. National survey data collected in the validation of an earlier version of the APL Survey, the Texas APL test, were used to determine the validity of the latter test.

It was found that the Texas APL test had concurrent validity because of high positive correlations with measures of success in adult life but that the test lacked construct validity because it measured three distinct skills, verbal ability, writing, and computation rather than the five skills and five content areas claimed by the test developers. Also found was that the APL Survey has neither concurrent validity because of low correlations with measures of success nor construct validity because it measures only one skill, reading, rather than five skills and five content areas. The scores on the APL Survey and the GED test were highly related ($r = .81$) and it was concluded that this was due to the heavy stress on reading skills on both tests.

The APL and GED approaches to measuring outcomes are not substantially dissimilar and high school diplomas issued on the basis of these two testing approaches would have about the same meaning. The value of introducing the APL approach into ABE is questionable on the bases that the APL tests may not measure what their developers claim, and that if the APL approach is not substantially different from the GED approach the costs of integrating the former approach into ABE may exceed the expected benefits of doing so.

AN EXAMINATION OF THE VALIDITY AND UTILITY OF A LONG WORD MEASURE IN THE PREDICTION OF READABILITY Order No. 7926110

CRAMER, Genevieve Rummell, Ph.D. University of Minnesota, 1979. 161pp.

The purpose of the study was to evaluate the validity and utility of a new long word measure of vocabulary difficulty and the readability formula employing that measure using three kinds of analyses. The new measure (proportion of words of six or more letters) and the Raygor readability estimate based on it were evaluated using comparative validation, validation against external criteria of text difficulty, and comparisons of formula application rates.

First, the Raygor estimate was compared with six established formulas--the Fry, Dale-Chall, Smog, Fog, Spache, and Flesch Reading Ease formulas--correlating grade level designations assigned by the formulas on 165 hundred-word passages taken from SRA Reading Labs IIIa and IVa. A computer readability analysis program (MNIRAP, Issue, 1977) applied to the passages yielded grade level designations and specific predictor measure information for each passage. Pearson product-moment intercorrelations among predictor variables were computed and grade placement intercorrelations among formulas were compared. The Raygor estimate scores correlated highly with those of the Flesch (.936), Fry (.901), Fog (.898), Smog (.898), Dale-Chall (.850), and Spache (.599). Similar significant coefficients were found when a second set of correlations were computed making comparisons on a common universe of data. Complete data were available on 106 of the original 165 passages. The results of six stepwise multiple regression analyses compared the relative weights of seven predictor variables used in determining grade placements for the 106 passages.

The second part of the study determined the extent to which five reading selections predicted to be at different grade levels (6th, 7th, 8th, 10th, and 11th) by the Raygor estimate reflected those levels when subjected to four external criteria of difficulty. Subjects were thirty college students enrolled in a developmental reading course at Southwest Missouri State University. Subjects were randomly assigned all five selections--found in How To Become A Better Reader (Witty, 1962)--with the constraint that each selection was assigned to an equal number of students during the five administrations. Analysis of variance results for the four dependent measures--reading rate, multiple-choice comprehension test results, multiple-choice vocabulary test results, and readers' judgments--showed the means differed significantly among the five selections. No significant main effect of order or interaction of order and difficulty level was found. Post hoc analyses using the Newman-Keuls procedure indicated only the selection predicted to be most difficult differed significantly from the remaining selections.

Finally, to test the contention that the Raygor estimate is quicker to apply than the widely-used Fry readability graph, thirty-seven analysts were randomly assigned to apply either the Fry or Raygor procedure to a set of three hundred-word passages. Subjects were college students taking a pre-service secondary teacher training course. After applying the first formula under timed conditions, the analysts applied the alternate formula to a slightly altered version of the passages. A correlated t-test showed a significant difference ($p < .001$) between the mean application rates using the two formulas. The application rate mean for the Raygor estimate was 12.77 minutes and for the Fry was 16.61 minutes.

Two of the three parts of the study support usage of the long word measure and the Raygor estimate. The comparative study data suggest the long word measure weighs elements similar to those employed in the other formulas. Intercorrelations among the formulas indicate the Raygor estimate was consistent with the other formulas in readability prediction. The comparison of analysts' application rates indicates the Raygor is quicker to apply than the Fry graph. Further research is needed before an evaluation can be made of the effectiveness of the Raygor estimate in relation to external criteria of text difficulty.

AN INVESTIGATION OF SPEECH RATINGS AS A FUNCTION OF RATING SCALE FORM, EVALUATOR TYPE, AND TRIALS Order No. 7924414

EDMUNDS, Robert French, Ph.D. Ohio University, 1979. 92pp. Director of Dissertation: Dr. Richard F. Whitman

The problem of classroom speech evaluation has been a topic of the literature for the past fifty years. Reliable quantitative measures have been advocated as well as more general types of assessments of student delivered classroom speeches. The purpose of this study is to test the feasibility of utilizing three methods of evaluating student delivered classroom speeches (peer-group only, peer-group/instructor, and instructor only) using two types of evaluation forms (composite rating form and general-effectiveness rating form) over a period of time/trials. This study is based in part on a study conducted by Wiseman and Barker in 1965-66 at Ohio Univer-

sity in peer-group instruction. The present study contains three areas of investigation. The first area is an attempt to determine if there is a difference among peer-group only, peer-group/instructor, and instructor only evaluation conditions. The second part of the investigation is an attempt to discover whether there is a difference between raters (peers and/or instructors) when using either a composite of ten rating scales or a single general-effectiveness rating scale. The third area of investigation is to determine if there is a difference between the scoring of student delivered classroom speeches among evaluating methods and rating forms over a period of time/trials.

The subjects for this study were students enrolled in the first course of public speaking. The instructors of those classes also served as subjects for the study. The six experimental conditions were assigned on a random basis since the experimenter had no control over the registration procedures. A rating form was constructed for the study containing six rating scales designed to measure content and four rating scales designed to measure delivery. A single general-effectiveness rating scale was used as the second test instrument. The data consisted of the results of the ratings of each of four speeches given by the members of each of the classes included in the study. The above concepts led to the formulation of two hypotheses, stated in the null form, to determine if there was a difference among three different groups of raters using specified rating forms. A third hypothesis was formulated to determine if there was a difference between two types of rating forms, a composite rating form or a general-effectiveness rating form. A fourth hypothesis was formulated to determine if there was a difference based upon time/trials between the groups and/or the types of rating forms used.

The results of the investigation point to three general conclusions. (1) There is no difference between peer-group and/or instructor evaluation when using either a composite rating form ($p > .05$) or a general-effectiveness rating form ($p > .05$). (2) There is a difference when peer-groups and/or instructors using either composite rating forms or general-effectiveness rating forms are compared ($p < .05$). (3) There is a difference over time/trials when peer-groups and/or instructors use different types of rating scales.

THE RELATIONSHIP BETWEEN HIGH SCHOOL READING AND MATH ACHIEVEMENT LEVELS AND LATER LIFE EXPERIENCES: IMPLICATIONS FOR MINIMAL COMPETENCY TESTING Order No. 7914348

FURR, Jane Dellana, Ph.D. The University of North Carolina at Chapel Hill, 1978. 1032pp. Supervisor: Dr. R. Neill Scott

This study investigated the relationship between high school reading and math achievement and later life experiences of average and below average achievement high school graduates. The life experiences variables included variables in the categories of work/employment, education/training, and personal/family. The purpose of the study was to provide information regarding the life experiences of average and below average high school graduates. This information should be useful to those governmental and educational personnel who compose competency testing rationale and objectives and establish the tests' passing levels.

The literature review revealed no studies which had been conducted regarding the life experiences of people who had taken minimal competency tests. Nor were any studies found which investigated the post high school experiences of low achievement high school graduates. Those making minimal competency decisions had little empirical information regarding the vocational, educational, or personal experiences of the low achievement high school graduates

The data sources for the study were the base year, first year and second year follow-up surveys of the National Longitudinal Study of the High School Class of 1972. The NLS Student Test Book reading and math achievement scores were used to determine the subjects' achievement levels for the reading, math, and read/math (the average of the reading and math scores) achievement variables. It should be noted that this test was not designed as a minimal competency test, but as an achievement test. The three achievement variables each had three levels: low, marginal, and average. The low achievement levels were made up of those subjects who, if they had been given a minimal competency test in high school, would probably not have passed the test. The marginal achievement levels were made up of those subjects who, if they had been given a competency test in high school, would probably have narrowly passed the test. The average achievement levels were made up of those subjects who, had they taken a competency test in high school, would have passed.

The basic research question was: What is the relationship between high school reading and math achievement levels and later life experiences? These life experiences included the following: (1) employment; (2) occupational category; (3) steadiness of employment; (4) job satisfaction; (5) perceived use of high school training on the job; (6) yearly income; (7) post secondary school attendance; (8) type of post secondary educational institution attended; (9) attainment of a license, certificate, or degree; (10) perception of high school curriculum; (11) marital status; (12) number of dependents; (13) voter registration and voter participation; (14) use of consumer strategies.

The analyses indicated that little relationship existed between high school reading and math achievement levels and later life experience variables of below average high school graduates. Additionally, it was found that: (1) single males had a greater yearly income than single females; and (2) in each achievement level, a greater percentage of males than females were employed.

Several reasons were offered for the finding of a lack of relationship between high school reading and math achievement levels and life experience variables. The study certainly indicated that further research is necessary in the area of basic skills to set the appropriate minimal competency testing objectives and passing levels. Without further research, young people may be labeled "incompetent" on the basis of the examination of tasks which may bear little relationship to later life experiences.

SAMPLE FREQUENCY MODIFICATIONS AND COMPARATIVE READABILITY OF FOUR READABILITY FORMULAS AS APPLIED TO SELECTED TEXTBOOKS ON THE TEACHING OF READING Order No. 7918451

GROVE, Martha Judin, Ph.D. East Texas State University, 1979. 148pp. Adviser: James B. Wilson

Purpose of the Study: The purpose of this study was to determine whether more efficient sampling procedures could be found for the Flesch Reading Ease Formula (Flesch) and Dale-Chall Readability Formula (Dale-Chall) and whether the Fry Readability Graph (Fry) and SMOG Grading Formula (SMOG) agreed with the Flesch and Dale-Chall formulas.

Procedure: The standard sampling procedures described by the authors of the formulas and three modified sampling procedures were tested on the four readability formulas, each of which was applied to twenty-one selected textbooks. Six hypotheses were constructed to test the differences among the standard and modified sampling procedures.

An analysis of variance was used to test for significant differences among the various sampling procedures. The Scheffé method of multiple comparison was used to make specific comparisons whenever the F reached significance. The Pearson product-moment correlation coefficient was used to test every possible relationship of the readability formulas and sampling procedures.

Findings:

1. There were no statistically significant differences among the mean of the standard sampling procedure and the means of the modified sampling procedures when applied to the Flesch, Dale-Chall, Fry, and SMOG formulas.

2. Using the standard sampling procedure for each formula, respectively, there was a statistically significant difference between the mean grade equivalent scores for the Flesch and Dale-Chall formulas and the Flesch and Fry formulas. Also, there was a statistically significant difference between the mean grade equivalent scores for the Dale-Chall and Fry formulas and the Dale-Chall and SMOG formulas.

3. Using the standard sampling procedures and the modified sampling procedures in all combinations, there were no statistically significant relationships among the four formulas in fifty-four of the ninety-one possible comparisons tested.

Conclusions:

1. Use of the modified sampling procedures tested with the Flesch and Dale-Chall formulas yielded a mean grade equivalent score as accurate as the standard sampling procedures and required substantially less time. Therefore, the classroom teacher can obtain sufficiently accurate results by using any of the more time-efficient, modified sampling procedures.

2. When each of the four readability formulas was tested with its standard sampling procedure, statistically significant differences were found. However, when the standard error allowed by each author is taken into consideration, the Fry and SMOG formulas appear to agree with the Flesch and Dale-Chall formulas. For practical classroom application, therefore, either the Fry or SMOG formula can be used with as much accuracy as the more time-consuming Flesch or Dale-Chall formula.

3. When the comprehension criteria applied to each formula by its author(s) are taken into consideration, the ranking of the grade equivalent scores for the formulas would be expected to be highest for the Dale-Chall with the Fry, Flesch, and SMOG progressively lower. However, this ranking was not found in this study.

4. In all cases, a statistically significant relationship was found when any of the modified sampling procedures was used with the readability formulas tested. Moreover, there was a statistically significant relationship found between modified sampling procedure B (six 100-word samples per book) and the standard sampling procedure for the Flesch, Dale-Chall, and Fry formulas, respectively. This finding suggests increased confidence in the use of the more time-efficient modified sampling procedure B by the classroom teacher when using either the Flesch or Dale-Chall formula.

THE DEVELOPMENT OF AN INSTRUMENT TO MEASURE HUMAN COMMUNICATION FUNCTIONAL DOMINANCE

Order No. 7927106

JOHNSON, Sally Hartman, Ph.D. University of Denver, 1979. 123pp.

The purpose of the present study was to develop an instrument to measure the construct "Human Communication Functional Dominance" based upon the three functions of human communication proposed by F.E.X. Dance and C.E. Larson. These functions are: (1) linking; the ability to respond to changes within the environment in an effort to reduce entropy via spoken symbolic adaptation or decentration processes; (2) mentation; the development of higher mental processes which includes memory, planning and foresight, intelligence, judgment, etc.; and (3) regulation of behavior; specifically, the verbal control of human behavior. Briefly, human communication functional dominance is the extent to which an individual relies upon behaviors characteristic of one of these functions of human communication over the others in various speech communication situations.

Chapter Two was devoted to developing the construct Human Communication Functional Dominance by describing in detail these three functions, specifying exhibited behaviors which represent these functions, and defining in greater detail human communication functional dominance.

Once the construct Human Communication Functional Dominance was outlined, the following general steps were taken to develop an instrument to measure that construct: (1) selection of an appropriate measurement methodology; (2) development of an item pool; and (3) refinement of the instrument following evaluation of the items by experts on the Dance and Larson functions, a pre-test, a constructivism check, and two statistical item analyses. The final instrument resulting from these steps included 24 situationally bound items. This instrument was then tested for validity and reliability.

Test-retest reliability using a sample of 30 subjects randomly selected from the basic Communication course at the University of Wisconsin-Milwaukee yielded an $r = .73$ for linking; $r = .68$ for mentation; and $r = .46$ for regulation.

Two forms of validity were considered to be important for this study; content and concurrent. Content validity was assessed by a panel of experts on the three functions. This panel determined whether the content in the descriptions of the three functions and the behaviors representing them as specified in Chapter Two were accurate and whether the items based upon these descriptions were appropriate. This review suggested that there was support for the content validity of the items.

The concurrent validity of the instrument was investigated by correlating subscores on the Human Communication Functional Dominance instrument with appropriate subscores of W. Schutz's Fundamental Interpersonal Relations Orientation - Behavior (FIRO-B). This resulted in the following correlations: linking vs. Expressed Inclusion $r = .43$; linking vs. Expressed Affection $r = .39$; regulation vs. Expressed Control $r = .24$. Correlations used to investigate concurrent validity using another instrument as the criterion are expected to be moderately high, but not too high, therefore, the correlations used to test the linking portion of the instrument were considered to be adequate. The mentation portion of the Human Communication Functional Dominance instrument was not investigated for concurrent validity because an appropriate instrument could not be identified.

THE ESTABLISHMENT OF RELEVANT SIXTH-GRADE NORMS ON A READING TEST CONSTRUCTED FOR GRADES 7-9

Order No. 7921246

LEIMGRUEBLER, Zoe Walker, Ph.D. The University of Oklahoma, 1979. 169pp. Major Professor: Dr. Robert Curry

This study was designed to establish local sixth-grade norms on the Gates-MacGinitie Reading Test, Survey E. This test was constructed and nationally normed in 1964 for grades 7-9, and regionally normed for superior-achieving students in grades 5-9 in the 70 districts comprising the Metropolitan School Study Council in New York (1965). The local subjects were students in Putnam City, a suburban Oklahoma City public school district in which grade level means of reading test scores are consistently found to be superior to both their grade placement and anticipated achievement. Pilot study findings (1973-76) indicated that the test ceiling for Survey E was appropriate for the local sixth-grade population. The need was therefore to provide relevant norm scales with which to interpret the reading test data required for annual placement of these students in ability-grouped seventh-grade reading classes.

The Gates-MacGinitie Reading Test, available in three parallel forms, is comprised of subtests to measure Speed and Accuracy, Vocabulary, and Comprehension. The three forms were randomly distributed to all sixth-grade students ($n = 1478$) exclusive of those enrolled in classes for the Educable and Trainable Mentally Retarded. The tests were administered in mid-February, 1977, by these students' own reading teachers in the normal classroom environment. Students recorded their responses on machine-scorable answer forms which were later coded to include the individual IQ score data resulting from the previous October's administration of the Short Form Test of Academic Aptitude.

Means and standard deviations of IQ scores were generated for each of the three norm samples. These statistics, as well as the standard error of measurement, the Kuder-Richardson Formula 20 (reliability), and norm scales expressed in percen-

tiles, percentile bands, z-scores, T-scores, and stanines were calculated for each form of each subtest. In addition, the Gulliksen and Guilford formulae for reliability of speeded tests were applied to the Speed and Accuracy test data.

No statistically significant differences were found among local sample means of IQ nor reading subtest raw scores (F-test for the Analysis of Variance; .05 level). The local sample means for each subtest and form were compared with (a) those of the seventh-grade portion of the original norm group (two-sample z-test for an hypothesis about a mean; .05 level); and with (b) the raw score equivalents of the interpolated grade equivalent score 6.5 on both the original and MSSC norm scales (one-sample z-test for an hypothesis about a mean; .05 level). In all instances, the local sample means were found to be significantly lower than those of other groups on the Speed and Accuracy test. Comparisons of local and seventh-grade sample means for Forms 1 and 3 of the Vocabulary and Comprehension subtests yielded findings of no statistically significant differences; however, the local group mean was superior on Form 2 of these tests. Comparisons of local means with the 6.5 raw score equivalents revealed the significantly superior performance of (a) the local group over the original sample on all forms of Vocabulary and Comprehension; and (b) the MSSC group over local samples on all forms of all subtests.

The conclusions reached were (a) that the local samples were equivalent both in academic aptitude and reading test performance; (b) that the three test forms can be used with equivalent results with this population; (c) that there was insufficient evidence to consider the Speed and Accuracy test a reliable instrument for use with this population; and (d) that, in support of findings in the literature, the use of local/special-study norms yields a more appropriate interpretation of raw scores for scheduling and placement than does the use of national norms with groups whose mean performance indicates that they cannot be considered a representative sample of the original norm group.

THE DEVELOPMENT AND VALIDATION OF THE PRE-SCHOOL LANGUAGE-COGNITIVE SKILLS ASSESSMENT FOR CURRICULUM ENTRY Order No. 7923972

LOFGREN, John O., Ed.D. University of the Pacific, 1978. 129pp. Chairman: Dr. Hugh J. McBride

PURPOSE: This study constructed an assessment instrument for the use of teachers and their assistants in preschool education programs. It also established measures of reliability and validity of that instrument.

PROCEDURES: Initially, extant assessment instruments were scrutinized. From ideas generated by the best of these instruments and from lists of competencies necessary for academic success in grade one of public schools, the first draft of the instrument was prepared. Extensive revision to reduce the length of the assessment and to simplify its administration was then accomplished. The second edition was then field-tested in preschool centers to ascertain whether (a) the instrument was economical in respect to time, (b) whether the language of the items would be comprehensible by the age levels of children examined, and (c) whether the interest level of the instrument would hold their attention.

Following field-testing minor revisions were made before submitting the assessment to three expert judges for item-by-item scrutiny to establish content validity. Judges considered (a) relevance of item content to curricula stressing language and cognitive skill development, (b) age-level placement of items, and (c) adequacy of item language in eliciting considered responses. From the critiques of the judges, further revisions to the assessment were made.

The test-retest reliability of the instrument was next determined by assessing a group of pupils twice each with a ten day interval separating pairs of assessments and then computing a Pearson product moment correlation for the pairs of scores. Next interrater reliability was determined using the Kendall coefficient of concordance technique upon scores obtained by each of four raters who assessed each of a group of ten pupils.

Criterion validity was then examined by applying the Spearman rank order positions of pupils according to assessments scores with rank positions assigned by their teachers judgments of their language and cognitive skills proficiencies. Teachers utilized a list of skills representative of assessment items for their guidance.

RESULTS: An assessment instrument was constructed sampling language and cognitive skill behaviors from age two through seven years, establishing a range to detect both slow and rapid development. The instrument was found to be (a) economical in terms of assessment time, (b) interesting to young children so that attention span poses no problem, (c) usable by preschool staff members who possess no psychometric expertise, (d) simple to administer and interpret, (e) possessing content validity, test-retest reliability, and interrater reliability.

RECOMMENDATIONS: Revisions of this instrument to enable accurate assessment of children whose primary language is other than Standard American English would be a contribution to educational practice in the United States. Mere translation, however, is not suggested. Also basic research to determine whether cognition precedes, occurs with, or follows language development would serve, heuristically, the development of preschool curricula. Predictive validity, investigated by a five year study in a community with population stability, would determine whether this instrument has value in preschool pedagogy.

A CORRELATION ANALYSIS OF LOCAL AND STATE MINIMUM COMPETENCY STANDARDS IN NEW JERSEY SCHOOLS Order No. 7914128

NADLER, Frederick Francis, Ed.D Rutgers University The State University of New Jersey (New Brunswick), 1978. 119pp. Chairperson: Dr. Frank H. Scherer

PURPOSE

The purpose of this study was to investigate the extent to which local and state minimum competency standards in reading and mathematics differ in the public schools in New Jersey. This study had also analyzed the relationships of local and state minimum competency standards in districts which vary in SES characteristics and in effort to support education.

PROCEDURE

Numbers of students below local and state standards in reading and mathematics at grades four, seven and ten were gathered from five hundred thirty-eight of the five hundred seventy-eight total school districts in New Jersey. The state standard was defined as the number of students falling below a sixty-five percent correct response level on the 1976 New Jersey Educational Assessment Program test. The local standard was defined as the number of students who fell below local minimum competency levels using locally selected tests or local teacher judgment. Local standard data were recorded from ESEA Title I applications for 1976-77. State standards data were obtained from the New Jersey State Department of Education (NJSDOE). Districts SES factors called District Factor Group (DFG) and district effort to support education factors called Cluster designation were obtained from the NJSDOE. The relationships of local and state standards were analyzed using the Pearson Product-Moment correlation technique.

RESULTS

The numbers of deficient students identified by local and state minimum competency standards were highly correlated at a significant ($p < .01$) level for both reading and mathematics at each grade level. The correlations within the low DFG districts were significantly ($p < .05$) higher than either the Middle or High DFG districts. Within High, Middle, and Low Cluster districts, local and state standards were significantly

correlated ($p < .01$) for reading at the fourth, seventh and tenth grade levels. The correlations within High Cluster districts were significantly ($p < .05$) lower than either the Middle or Low Cluster districts. The proportions of districts where the local standard exceeded the state standard by a five percent discrepancy were similarly distributed across the High, Middle and Low DFG Cluster districts. However, the tendency for local standards to exceed state standards was greater in the Middle DFG district when discrepancies of fifty, sixty-five and ninety percent between local and state standards were analyzed.

CONCLUSIONS

The high degree of correlation between local and state standards in reading and mathematics leads one to conclude that the relative needs of districts for remediation are consistent with both the local and state standards.

SES was found to be a variable in the correlations of local and state standards. Since correlations were significantly higher in the Low DFG districts, it might be interpreted that the state established the standard at or near the level established the standard at or near the level established by Low DFG districts.

Effort to support education was also found to be a variable in the correlation of local and state standards. These correlations were significantly lower for high effort districts. This may mean that the more effort there is to support education, the more likely the district is to set a standard which is discrepant from the state standard.

Local districts showed an overall tendency to set higher standards for themselves than did the state. Local and state standards were more likely to correspond in Low DFG districts. It might be interpreted that the state established the minimum competency standard at or near the level established by Low DFG districts.

The findings imply that statewide minimum competency standards really do represent a minimum level for the achievement of basic skills. Statewide standards do not seem to represent a maximum level for local districts.

TEACHER EXPECTANCY: THE EFFECT OF RACE, SEX, DIRECTION OF WRITING PERFORMANCE AND TRIALS ON THE GRADING OF ESSAYS Order No. 7919098

PORTER, Patricia Thomas, Ph.D. Texas A&M University, 1979. 104pp. Chairman: Dr. Glenn R. Johnson

This investigation was conducted to determine the effects of direction of writing performance, race, sex, and trials on teacher trainees' grading of students' essays.

The subjects for this investigation were 59 students, 51 female and 8 male, enrolled in English and Education classes. Subjects were assigned random numbers and subsequently placed into groups, designated Up, Down, and the Same. The essays to be graded by the Up group were progressively better over a period of five trials. The essays distributed to the Down group of subjects were progressively poorer over a five trial period. The essays distributed to the Same group were approximately the same. Rankings of the essays, as to which were better, which were poorer, and which were approximately the same in quality, were determined by ranking the mean scores given the student essays by a panel of experts in the Department of English at Texas A&M University.

A 4-way repeated measure analysis of variance revealed a significant F value on only one main effect, trials, $F(4,163) = 3.5$, $p < .05$. Other main effects had no significant influence on teacher trainee expectancy of students' academic ability, as evidenced through the grading of essays (direction of writing performance, $F(2,47) = .93$; race, $F(1,47) = .13$; sex, $F(1,47) = .39$). One interaction effect, direction of writing performance X trials, was found to have a significant effect on the expectancy of a student's academic ability, as evidenced by the teacher trainee's grading of the student's essays ($F(8,163) = 12.32$, $p < .05$).

PRE-EMPLOYMENT TEST SCORES AS PREDICTORS IN A REMEDIAL READING PROGRAM Order No. 7927500

RENTZ, Sarah Taylor, Ed.D. The Louisiana State University and Agricultural and Mechanical College, 1979. 80pp. Supervisor: Eric L. Thurston

Problem

The purpose of this study was to answer the question, "What is the predictive merit of various pre-employment tests in terms of time needed to upgrade reading skills to the apprentice level?"

Procedure

An upgrade program was designed to improve the reading level of low skilled workers to a point where they could be successful in an apprentice program. Comprehension, vocabulary, sequencing, and study skills were the basic reading skills taught in the pre-apprentice program.

Ninety-nine low skilled workers at a petro-chemical complex participated in the two year remedial reading program. These individuals had been employees of the industry for a number of years. The sample of this study consisted of eighty-four employees who met the terminal reading criterion.

Analysis of Data

A battery of tests were administered to the participants of this study before they were hired by the industrial complex. This battery included four standardized tests: 1) Bennett Mechanical Comprehension Test (ME), 2) Advanced California Mathematics Test (MA), 3) Test of Chemical Comprehension (CH), and 4) a learning abilities test called Personnel Questionnaire (LA). A coefficient of correlation was run between the pre-employment test scale scores and the length of time each employee required to meet the terminal criterion in reading.

Findings

In review of the five coefficients of correlation it was found that three of the variates (MA, ME and composite scores with hours) produced no significant value in predicting time required to upgrade reading skills to the apprentice level. Two of the correlations (LA and CH with hours) were significant at the .05 level.

The predictors (LA and CH) having sufficient validity were combined to obtain the highest multiple correlation (R) with the criterion. The Personnel Questionnaire and Test of Chemical Comprehension correlation scores of -.29 and -.24 and an intercorrelation of -.12 yielded a multiple R of .40 signifying an improvement in predictive efficiency over that from either test alone. This coefficient was significant at the .01 level.

A regression equation was derived to predict the training time from The Personnel Questionnaire and Test of Chemical Comprehension scores. From this equation a regression table was constructed to facilitate the anticipation of remedial reading times when scores from the LA and CH tests were known.

Conclusions

The results from this study indicated that the Bennett Mechanical Comprehension Test, Advanced California Mathematics Test and the composite test scores had no predictive merit. However, coefficients of correlation for the Chemical Comprehension Test and The Personnel Questionnaire were significant and warranted further investigation concerning predictive values.

The petro-chemical industry of this study could conclude that a prospective employee with a low CH score and a low LA score could be expected to require a longer length of time to upgrade reading skills to that of the apprentice level than an individual with a higher score on these two tests. The predictive equation constructed from this study could be expected to be accurate by using the standard error of estimate 2 in 3

instances. Factors such as physiological, psychological and sociological problems could be expected to affect the accuracy in 1 out of 3 situations.

A FACTORIAL ANALYSIS OF ESSAY EVALUATIONS

SHALE, Douglas Gordon, Ph.D. University of Toronto (Canada), 1978

Criteria were obtained from 19 department heads of English in various Ontario high schools for the marking of each of expository and descriptive prose. Consensus was obtained on two lists of criteria (one for each mode of writing) through a modified Delphi process. Seventeen criteria were found to be common to both modes of writing – 3 additional criteria were specific to exposition and 2 to description. Four markers applied the appropriate set of criteria to 168 expository and 168 descriptive essays written by 168 grade 12 students from various high schools throughout Ontario (each student wrote an expository and a descriptive essay). Intercorrelations were calculated among the criteria in both sets using data pooled over markers. The factor structures of these two correlation matrices were investigated by the common factor analysis model, by calculating the number of units of variance accounted for by various factors over various factor solutions and by calculating congruence coefficients for selected pairs of factors. The generalizability of factors over these two data sets was investigated qualitatively by interpreting the factors obtained from them, and quantitatively through covariance structures analysis and by calculating congruence coefficients for appropriate pairs of factors.

Exploratory factor analyses of the two intercorrelation matrices indicated that clusters of variables characterizing the factors remained quite stable as the number of factors prescribed changed. Four stable factor groupings emerged from the evaluations of both modes of writing. Two of the factors found in the description results corresponded closely to two factors that emerged from the exposition results. One of these was a "cohesiveness at the level of the paragraph" factor, and the other was a "mechanics as they render the substance of the essay ambiguous or unambiguous" factor. A third factor found for each of the essay types contained a number of criteria in common – namely, content, sentence variety, originality, and literary devices. However, in the case of description, the comparatively high loadings of descriptive detail and dominant impression suggest that the factor may emerge because these criteria have been applied to descriptive prose. In the case of exposition, content and facts and illustrations loaded highest on this factor, and this suggests that so far as exposition is concerned, the factor is better interpreted as a "content" or "substance of the essay" factor. The fourth exposition factor seems to be a mode specific factor because of the high loadings of the criteria, thesis and conclusion on this factor. The fourth description factor may be described as a "cohesiveness at the level of the essay" factor because the criteria form, coherence, point of view, and emphasis in the whole essay load highest on it. Three of the congruence coefficient values calculated among the appropriate pairs of factors were very large, further indicating that three factors are very similar over the two types of writing. The residuals derived from the various factor solutions were generally uniformly small and provided additional evidence that these factor groupings were reasonable representations of the data.

Because no statistically significant representation of the data sets for a (small) specified number of factors resulted, it was not appropriate to apply a covariance structures model to test the hypothesis that the 3 factors in common were identical (in the sense implied in the model). A test of the hypothesis that the common parts of the two correlation matrices were the same (also in the covariance structures sense) was inconclusive because of the magnitude of the amount of computing required.

Mark/re-mark correlations were calculated for a sub-set of 15 essays of each type. These correlations were, in general, quite low indicating that this rendering of the reliability of assessments derived from the criteria was poor. These low correlations might be regarded as an indication that little additional 'reliability' may be realized in essay grading through a highly analytic marking scheme.

There was a substantial disparity between how the criteria were perceived (as indicated from a Free-sort procedure) and how they were applied (as indicated by the factor groupings). This disparity may be an indication that the markers were unable to apply the criteria in a manner congruent with their conscious perceptions of them, in which case the disparity would represent a serious limitation to the extent to which precision of marking can be realized through an analytic marking scheme. However, before such a conclusion is accepted, more information needs to be obtained regarding the comparability of groupings of criteria obtained in the abstract with no reference to specific written composition and groupings derived from applying the criteria to the evaluation of written work.

A DESCRIPTIVE STUDY OF COMPETENCY BASED EDUCATION AS RELATED TO ACHIEVEMENT, READING ABILITY AND GRADE POINT AVERAGE OF HIGH SCHOOL STUDENTS

Order No. 7926582

SMITH, Carolyn Crawford, Ph.D. Kansas State University, 1979. 108pp.

The purpose of the study was to test whether a pilot program for competency testing provided results that correlated with other measures of academic performance. The test was given to seniors at Sardis Town High School in Mississippi and was developed with funds provided by the federal government under Title IV C. Scores were compared with results from the Test of Academic Progress, the reading component of that examination, overall grade point average of students, rank in class and whether the student was college bound. The study also tested whether scores were related to sex and race of students.

The sample population included 120 seniors at the predominantly Black, rural and low income school. Of the total, 108 students were Black and 12 were White. Sixty-nine were female and 51 were male. Information about scores, class ranking, race and sex were obtained from personal files of the students.

The major findings of the study were:

1. The competency test was a weak predictor of scores on the test of academic progress and on the reading component of the test of academic progress.
2. There was no significant relationship between scores on the competency test and the overall grade point average, class rank, and future educational plans.
3. Female students scored significantly higher on the competency test than male students.
4. There was no significant relationship between race and performance on the competency test.
5. There was no significant relationship between academic achievement and reading achievement.

The major conclusion of the study was that the test did not measure features of the curriculum that were taught and evaluated at Sardis Town High School. The failure to test for items of the curriculum raised potential due process issues. This would be the case if the test were found to be arbitrary and capricious as a measure of student performance. The study suggested further research before the test is used on a statewide basis.

The following major recommendations were made:

1. The conducting of additional research using larger sample populations from diverse social settings.
2. The conducting of further study of the relationship between race and test results. The samples should be larger and there should be a greater balance between the number of Black and White subjects.
3. As a preliminary to developing a means of weighing test results to eliminate the relative advantage of females, study of possible sex bias in test items is recommended.
4. The initiation of dialogue within the profession and between professionals and the public on the merits and costs of competency testing is a final major recommendation.

THE USE OF KAUFMAN'S FACTOR ANALYSIS OF THE WISC-R IN DISTINGUISHING ABLE AND DISABLED READERS AT THREE GRADE LEVELS Order No. 7928237

SMITH, Merrill Lee, Ed.D. Oklahoma State University, 1979. 84pp.

Scope of Study: The specific purpose of this study was to statistically distinguish between the two types of reader, able and disabled, at three levels of development: grade two, grade four, and grade six. Three discriminating variables, Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility were chosen as characteristics on which the groups were expected to differ. These three variables were derived from the Wechsler Intelligence Scale for Children-Revised. Four hundred sixty-six elementary school children were screened with the Lorge-Thorndike Non-Verbal Intelligence Test and the Gates-MacGinnitie Silent Reading Comprehension Tests. From this number, samples were drawn which consisted of 60 pupils at each of three different grade levels: a group of thirty disabled readers at grades two, four, and six, and a group of thirty able readers at second, fourth, and sixth grade. In all, 180 elementary students were used in this investigation. Disabled readers are defined as: second graders achieving .5 of a year or more below expectancy, fourth graders achieving .8 of a year or more below expectancy, and sixth graders achieving 1.2 years or more below expectancy. Able readers are defined as those readers achieving at or above their expectancy level. Expectancy was determined by the years in school method developed by Bond and Tinker. The subjects for this study were selected from four schools in north-central Oklahoma.

Statistical Analysis: Independent samples t-tests were conducted to test the difference between the means of able and disabled readers on Verbal Comprehension, Perceptual Organization, and Freedom from Distractibility. Three, within-grade, two-group discriminant function analyses were conducted to determine whether or not group membership could be predicted on the basis of the three Kaufman factors.

Findings and Conclusions: The results of this study showed that the information possessed by the Freedom from Distractibility factor is most important in distinguishing between able and disabled readers at grade two. Its importance is nearly one-third greater than knowledge of Verbal Comprehension, and Perceptual Organization contributes no information to group separation at grade two. At grade four, Verbal Comprehension contributed most to group separation. Freedom from Distractibility contributed less than one-half as much information, and Perceptual Organization contributed no useful information in distinguishing between the groups at grade four. All three factors contributed to group separation at grade six. Verbal Comprehension was the predominant discriminator with Freedom from Distractibility and Perceptual Organization contributing less, but near equal amounts of information, at grade six. These factors, based on students of known group membership, distinguished between the groups with 73 per cent accuracy at grades two and six, and with 75 per cent accuracy at grade four.

The results of this study indicate that identification of the distractibility factor at grade two may be just as important as identification of an appropriate methodology for the child. While children tend to outgrow their distractibility, the cumulative effects of academic failure make early recognition of this deficit crucial to later development in reading. Support was not found in this study for perceptual-deficit explanations of reading disability at grade two, while, beginning at grade four, verbal-language deficiencies may explain reading disability.

A COMPARISON OF ERRORS MADE ON THE WEEPT AND THOSE FOUND IN STUDENT WRITING SAMPLES AS BASES FOR PLACEMENT IN FRESHMAN COMPOSITION CLASSES Order No. 7922196

STENBERG, Daniel Raymond, Ph.D. Indiana University of Pennsylvania, 1979. 268pp. Chairman: Marilyn S. Sternglass

This descriptive study was conducted using the two part English placement test given to entering freshmen at Southwestern Michigan College, Dowagiac, Michigan, during the 1977-78 college year. The study reviewed and categorized the forty item objective test, the Written English Expression Placement Test (WEEPT) and compared findings with items categorized on a Student Sample Essay. The overall study purpose was to determine whether the WEEPT alone could place students in a freshman composition course in which texts, syllabi and level of instruction would best suit tested needs or whether the WEEPT best served when used in conjunction with a student sample essay. A second purpose was to place the study within current concerns about objective versus essay testing as used for placement and to provide an analysis useful to departments of English in open door admission colleges. Included are a detailed explanation of the categories of the WEEPT along with the specifics of the grammatical structures tested by those categories, an explanation and specifics of additional categorized items found on Student Sample Essays but not covered by the WEEPT, and an analysis of individual items found in Student Sample Essays neither covered by the WEEPT nor in sufficient number to categorize but still seen as items of use in placement.

Following an analysis of WEEPT questions and the formation of WEEPT categories, twenty papers were randomly drawn from the thirty-five study period students who had completed only the Student Sample Essay. These papers were used to form additional categories not covered by the WEEPT but viewed as items of use in placement. Then, seventy-five tests were randomly drawn from the 135 entering freshmen who had completed both the WEEPT and the Student Sample Essay. The seventy-five were analyzed for WEEPT results and Student Sample Essay results with findings charted in terms of WEEPT categories, additional categorized items not covered by WEEPT and individual items found in Student Sample Essays neither covered by the WEEPT nor in sufficient number to categorize but still seen as items of use in placement. A comparison between WEEPT category findings and WEEPT category items generated in Student Sample Essays revealed that in four of the twelve categories the WEEPT was as or more discriminating in its results as validated, using t-tests and an analysis of variance with a significance at the .05 level of confidence. A comparison between numbers of errors committed in additional categories not tested by the WEEPT with WEEPT category errors matching both to syllabi, texts, and levels of instruction revealed that all five additional categories represented as large or larger percentages of error frequency than did nine of the twelve WEEPT categories and that three of the five additional categories best matched syllabi, texts and instruction at the remedial level, the most crucial to proper placement. Of thirty-two items not covered by WEEPT or additional categories and not found in sufficient number to categorize, twenty-two were revealed to be of use in placement. It was found that the WEEPT could best serve to place students in a freshman composition course matched to texts, syllabi and level of instruction when used in conjunction with a Student Sample Essay.

The study raises questions yet to be answered on future studies concerning qualitative assessment and matching tests, texts, syllabi, and instruction as well as the need to include spelling, punctuation and dialect matters in future quantitative and qualitative analysis of testing for placement purposes. The study concludes by noting its place in on-going concerns relative to placement and testing.

THE CONSTRUCTION AND VALIDATION OF A PROCESS MODEL FOR DEVELOPING INFORMAL ORAL READING INVENTORIES Order No. 7914103

WATKINS, William C., Ph.D. The Fielding Institute, 1974. 157pp.

Purpose. Develop a model for constructing informal oral reading inventories and establish some measure of the Model's validity.

Following careful review of the literature, a four step model that made extensive use of a readability formula was developed and field tested. Several classroom teachers developed one informal oral reading inventory by following the steps described in the Model. The informal oral reading inventory and selected standardized tests were administered to a sample size of 30 students from each of five grade levels (grades 2 through 6). The test scores were analyzed with selected statistical procedures.

Findings. An informal oral reading inventory construction model could be developed, and instructions for replicating the Model could be set forth in clear and concise language that teachers with varying years of experience could accurately follow.

Test scores yielded by the experimental informal oral reading inventory were compared with standardized test scores of the same subjects and four hypotheses were tested. Mean scores of the instruments were not equal except in two instances. However, grade equivalent scores fell within a single grade level and within the confidence intervals set for each test. Further, teachers' opinions substantiated the empirical validity of the informal oral reading inventory developed from the Model: teachers stated the informal oral reading inventory accurately predicted the instructional reading level of students.

Very positive correlations were found between and among the experimental inventory scores and the standardized test scores. The probability of the phenomenon occurring by chance was 5 in 10,000 times. The correlations for grades two, three, five, and six were all above .900. The correlation was .871 for fourth grade students.

Conclusions: (1) The literature attests to the importance of reading as a subject area. (2) Accountability and humanistic education are compatible and accomodate and facilitate better diagnostic, prescriptive, and individualized programs. (3) Some readability formulas have charts and graphs that can be applied easily to a variety of reading materials. (4) Oral reading tests and inventories are more alike than they are different, with at least ten common features identified. (5) Reading selections should be chosen from a set of basal reading textbooks, and the trend is to select passages from basal readers being used for the instructional reading program. (6) Essentially only one model for constructing oral reading inventories has been available, but the evidence reported in this paper strongly suggests that the Model used in this study is a valid and reliable alternative model. (7) Field testing procedures showed the Model used in this study is viable and replicable. (8) The sensitivity and validity of the Fry Readability Formula was reaffirmed. (9) Many stories and passages presented in a basal reader will have reading difficulty levels below and above the designated grade level assigned to the book. (10) The informal oral reading tests yielded scores that correlate positively.

Recommendations: (1) The Model should be replicated so its validity and reliability can be further verified. (2) Substitute other readability formulas and conduct similar studies. (3) The scores of informal oral reading inventories should be compared with scores of other standardized oral reading tests. (4) Test scores of students should be verified by students' classroom performance. (5) Use the model and construct an inventory for grades seven through twelve. (6) The Model, with some modifications, should be used to construct criterion referenced tests other than oral reading inventories.

Copies of the dissertations may be obtained by addressing your request to:

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106

or by telephoning (toll-free) 1-800-521-3042